

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
15 January 2004 (15.01.2004)

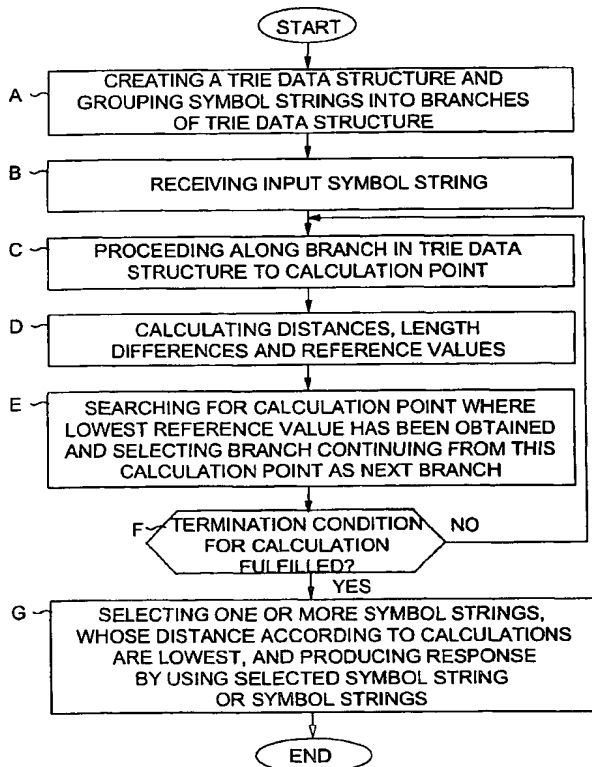
PCT

(10) International Publication Number
WO 2004/006126 A1

- (51) International Patent Classification⁷: **G06F 17/30** (74) Agent: **KOLSTER OY AB**; Iso Roobertinkatu 23, P.O. Box 148, FIN-00121 Helsinki (FI).
- (21) International Application Number: **PCT/FI2003/000540** (81) Designated States (*national*): AE, AG, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, DZ, EC, EE (utility model), EE, ES, FI (utility model), FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK (utility model), SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (22) International Filing Date: 3 July 2003 (03.07.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 20021330 5 July 2002 (05.07.2002) FI
- (71) Applicant (*for all designated States except US*): **SYS-LORE OY** [FI/FI]; Eerikinkatu 42, FIN-00180 Helsinki (FI).
- (72) Inventor; and
- (75) Inventor/Applicant (*for US only*): **HYVÖNEN, Jörkki** [FI/FI]; Laivapojankuja 3 b 38, FIN-00180 Helsinki (FI).
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO,

[Continued on next page]

(54) Title: SEARCHING FOR SYMBOL STRING



(57) Abstract: The present invention relates to a method for searching for an input symbol string, comprising receiving (B) an input symbol string, proceeding (C) in a trie data structure to a calculation point indicated by the next symbol, calculating (D) distances at the calculation point, selecting (E) repeatedly the next branch to follow (C) to the next calculation point to repeat said calculation (D). After the calculation (G), selecting the symbol string having the shortest distance to the input symbol string on the basis of the performed calculations. To minimize the number of calculations, not only the distances are calculated (D) at the calculation points, but also the smallest possible length difference corresponding to each distance, and on the basis of each distance and corresponding length difference a reference value is calculated, and the branch is selected (E) in such a manner that next the routine proceeds from the calculation point producing the lowest reference value.

WO 2004/006126 A1



SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations* AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG,

ZM, ZW), *Eurasian patent* (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), *European patent* (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), *OAPI patent* (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

- *of inventorship (Rule 4.17(iv)) for US only*

Published:

- *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/FI 03/00540

A. CLASSIFICATION OF SUBJECT MATTER

IPC7: G06F 17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC7: G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-INTERNAL, WPI DATA

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 1052576 A2 (MATSUSHITA ELECTRIC IND CO, LTD), 15 November 2000 (15.11.00), page 2, line 5 - line 8; page 3, line 19 - line 25; page 3, line 30 - page 4, line 55, page 5, line 50 - page 11, line 11, claims 1-3, abstract	1,5-7
Y	--	2-4
X	US 5377281 A (BALLARD, B.W. ET AL), 27 December 1994 (27.12.94), column 7, line 27 - line 36; column 7, line 60 - column 8, line 52, abstract	1,5-7
	--	

☒ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

Date of mailing of the international search report

30 Sept. 2003

06-10-2003

Name and mailing address of the ISA/
Swedish Patent Office

Box 5055, S-102 42 STOCKHOLM

Facsimile No. +46 8 666 02 86

Authorized officer

Åsa Viken /LR

Telephone No. +46 8 782 25 00

INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI 03/00540

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	SHANG, H. et al.: Tries for Approximate String Matching. IEEE Transactions on Knowledge and Data Engineering, vol.8, no.4, August 1996. See pages 540-547, figure 5 and abstract --	2-4
A	US 6377945 B1 (RISVIK, K.M.), 23 April 2002 (23.04.02), column 3, line 15 - column 4, line 63; column 7, line 12 - line 49; column 9, line 34 - column 10, line 21, column 11, line 64 - column 12, line 3, abstract --	1-7
A	UKKONEN, Esko: Finding Approximate Patterns in Strings. Journal of Algorithms 6, 132-137 (1985). See page 132-137 and abstract -- -----	1-7

INTERNATIONAL SEARCH REPORT

Information on patent family members

06/09/03

International application No.

PCT/FI 03/00540

Patent document cited in search report			Publication date	Patent family member(s)	Publication date
EP	1052576	A2	15/11/00	NONE	
US	5377281	A	27/12/94	NONE	
US	6377945	B1	23/04/02	AT 212736 T	15/02/02
				AU 4937099 A	01/02/00
				BR 9912015 A	10/04/01
				CA 2337079 A	20/01/00
				CN 1317114 T	10/10/01
				DE 69900854 D,T	22/08/02
				DK 1095326 T	13/05/02
				EP 1095326 A,B	02/05/01
				SE 1095326 T3	
				ES 2173752 T	16/10/02
				HU 0201630 A	28/08/02
				IL 140606 D	00/00/00
				JP 2002520712 T	09/07/02
				NO 311657 B	27/12/01
				NO 983175 A	11/01/00
				NO 993413 A	11/01/00
				PL 345714 A	02/01/02
				PT 1095326 T	31/07/02
				WO 0003315 A	20/01/00